Response to Comments
Draft Radiological Data Evaluation Plan Tech Memo
Former Hunters Point Naval Shipyard, San Francisco, California

PREPARED FOR:	HPNS Radiological Tiger Team
PREPARED BY:	CH2M HILL, Inc. (CH2M), Cabrera Services, Perma-Fix Environmental Services, Inc.
DATE:	January 26, 2017, updated February 24, 2017

The purpose of this document is to address comments on the Draft Radiological Data Evaluation Plan Tech Memo, Former Hunters Point Naval Shipyard. Comments were provided by Environmental Protection Agency (EPA), Department of Toxic Substances Control (DTSC), California Department of Public Health (CDPH), and Water Board. The Navy requested clarification on several EPA comments below in *italics,* followed by EPA clarification, followed by any additional comments on the responses. All responses to comments are provided in **bold**. The Radiological Data Evaluation Plan Tech Memo has been updated to address these comments. The plan and data quality objectives (DQOs) are intended to be part of a living data evaluation process and updated as additional information is obtained.

EPA Comments (January 11 and 17, 2017 and February 7 and 14, 2017)

General Comments

1. The quality of prior data has varied due to differences in analytical methods and other reasons. Historical data should be screened carefully for reliability and to ensure that any comparisons are meaningful and consistent prior to spending significant time and resources on statistical analyses.

   Data quality has been assessed during previous document reviews and approvals. The objective of this project is to identify manipulation of data that have already been deemed reliable with the exception of possible falsification. Appropriate statistical tests have been selected to assure comparisons and analyses are meaningful and consistent.

2. Where data quality is reliable, statistical tests are indeed another tool to find new evidence of potential falsification in previously undiscovered anomalies. However, prior data may also be unreliable because prior potential falsification has unknown scope. For example, earlier investigation and resampling has verified falsification of soil confirmation sampling. Other allegations have surfaced about potential disappearing of samples and tampering with data in spreadsheets. Because of unknown scope of prior falsification and the above challenges in data quality, statistical tests should not be used to automatically and definitively rule out locations to collect new samples in Phase 2. Sampling in many areas will be necessary regardless of the findings of the statistical tests.

   Navy Response: It is not clear what rationale is being suggested to sample in "many areas regardless of the findings of statistical tests."

EPA Clarification: EPA already made written recommendations October 28, 2016, for potential categories where sampling will be useful to improve confidence regarding actual health risk. These include (1) some of the areas of allegations from former workers and (2) areas of relatively higher risk prior to remediation. From recent conference calls, we understand that the Navy's independent consultants are developing more specific criteria for prioritizing areas to sample. EPA will provide input through the workgroup process to refine a plan. It is possible that as part of the emerging plan, some locations may be associated (1) or (2) above, and that, after statistical tests on data from those same locations, no anomalies are found. If such a situation arises, EPA may, on a case-by-case basis, still recommend that sampling should be prioritized if another reason for sampling exists (such as (1) or (2)), even if no anomalies are found, if the conclusions of a given specific statistical test are not reliable (e.g. due to limits related to the test itself or concerns about comparability of data or potential falsification of data). The EPA specific comments give examples of the types of potential concerns. We will work together with the Navy to refine the sampling plan through the workgroup process already established.

A generic UFP-SAP is currently being designed in parallel to the statistical data evaluations followed by addenda for location/task specific sampling as identified based on allegations, concerns, and the results of statistical tests and data evaluation. All plans will be provided to the Tiger Team for review and approval. The Navy agrees that it is possible that statistical tests may not identify all areas where alleged manipulation took place as reported by former workers. If such a situation arises, the Navy will work with EPA on a case-by-case basis to identify sites for additional sampling.

3. Due to the above concerns, random confirmation sampling should be performed in addition to investigating outliers identified as a result of the Data Evaluation Plan and compared with previous results.

   Navy Response: Please clarify the rationale for additional sampling if no evidence of falsification is found.

   EPA Clarification: In addition to the response above to General Comment #2, as stated in the original comments, given the unknown scope of potential prior falsification, then results of statistical analysis, specific allegations, and other methods of refining the scope of concern may not alone be always sufficient justification to not sample. However, EPA does agree that earlier priority for sampling should be in areas with a specific reason to have a higher expectation of prior falsification, due to specific allegations, findings of anomalies, higher risk prior to remediation, or some other specific reason. One category of random sampling could include sampling in randomly determined locations within the boundaries of areas of higher risk prior to remediation. We will work together with the Navy to refine the sampling plan through the workgroup process already established.

   As noted on a prior conference call by CH2M Hill, the Draft Data Evaluation Plan "May not find all areas where data manipulation has occurred."

   Additional "targeted location" but random sampling and analyses within targeted areas is recommended to confirm radionuclide concentrations and test for outliers with statistical

comparisons of Final Status Survey Results to reassure regulators, the general public, and to test performance of the statistical tests.

See response to EPA general comment #2. Sampling at an additional site that follows a "sampling approach" identified in the DQOs will be addressed with a Task Specific Plan (TSP); or if the sampling approach is new, the approach will be included as an addendum to the UFP-SAP, and the site will be addressed through a TSP.

4. Phase 1 and Phase 2 should be iterative. Sampling does not have to wait until full statistical analysis. Findings from field sampling should help inform statistical analysis.

Phases 1 and 2 have been and will continue to be conducted in an iterative fashion and results from field sampling will have statistical analysis conducted. See response to EPA general comment #2.

Specific Comments

1. The Draft Radiological Data Evaluation Plan, Former Hunter's Point Naval Shipyard (HPNS), San Francisco, California (the Plan) does not propose conducting a data quality evaluation. The Phase I Data Evaluation proposes to organize data sets into groups for each survey unit/trench unit or work area, parcel, and survey type to run statistical evaluations. Once the data is grouped, it is recommended that a data quality evaluation be conducted prior to statistical data analysis to ensure data sets are sufficiently comparable (for comparative analyses) and are of sufficient quality for decision making. Such a data quality review may include, but is not limited to, the following:
    a. Review of data sets to determine if a data validation was performed on the data
    b. If data validation was not performed, ensuring each data set was generated under a defined quality assurance/quality control (QA/QC) program, and includes results of QC samples
    c. Determining if detection limits (especially for Radium-226 [Ra-226] quantitated from gamma spectrometry analysis using the 186 kiloelectron volt [KeV] energy line) met the project requirements
    d. For comparison evaluations, ensuring the data sets being compared were generated using the same analytical methods, data reduction methods, and detection limits.

    This data quality review will ensure the statistical analyses are conducted on data sets with sufficient documentation regarding how the data was generated and are considered usable or comparable data so that the results of such statistical evaluations provide relevant and useful information. For efficiency, where data quality is not reliable or comparable, additional analysis may not be necessary or helpful.

    Navy Response: Items a, b, and c are beyond the scope of this project. The objective of this project is to identify manipulation of data. Assuring that detection limits met project requirements would have been part of the initial evaluation and report.

    EPA Clarification: The purpose of this comment is not to create new requirements for data reviews in and of itself. Rather, it is to encourage efficiency through preventing unnecessary work in situations where data may not be reliable or comparable. For example, in October, 2016, EPA recommended comparing parent and progeny results within decay chains as another

potential indicator of falsification. We appreciate the Navy has helped us understand the facts that in some situations, especially before 2010, some data may not be comparable due to differences in analytical methods and some data quality limits; therefore, any conclusions drawn from such a comparison would not achieve the intended result of identifying potential falsification. EPA would therefore no longer recommend that the Navy do comparisons in such situations. As another example, in areas where the error bars are large and the value is near the detection limit, a statistical test may not be a productive use of limited resources, since any conclusion drawn from such a test may not be reliable. Again, the recommendation for data quality review is for the purpose of evaluating the usefulness of statistical comparisons, not for the purpose of providing a data quality related basis for questioning prior cleanup approvals. We will work together with the Navy to refine the sampling plan through the workgroup process already established.

The Navy understands that parent and progeny analyses are best conducted on samples with a full 21-day ingrowth. See also responses to EPA general comments #1 and 2.

2. Discussions held during the Navy Scoping conference call on January 5, 2017, indicated that much of the historical laboratory gamma spectrometry data from the on-site lab reported results for Ra-226 that was quantitated using the 186 KeV energy line, rather than using the ingrowth method and quantitating Ra-226 from the daughter radionuclide Lead-214 (Pb-214) at the 609 KeV energy line. Quantitation of Ra-226 using the 186 KeV energy line is subject to interference from Uranium 235 (U-235) gamma ray emissions that occur at approximately the same energy and may be skewed due to higher amounts of backscatter photons in this energy range. Ra-226 data generated using the 186 KeV energy line may produce false positive results and may or may not be useable for decision making. Please clarify how Ra-226 data generated using the 186 KeV energy line will be handled.

   See response to EPA general comment #1. Ra-226 data is not treated differently based on how it was generated. If differences are identified, evaluations and investigations will identify the source of the differences.

3. The Phase 1 investigation should include some confirmation sampling of soil columns to test that the hypothesis of the statistical test used for Phase I for identifying outliers are valid. It would not be prudent to wait until after all of the Phase 1 and Phase 2 data reviews to begin confirmation sampling.

   Navy Reponses: It is not clear what is meant by "soil columns." Statistical tests were validated by ensuring they would identify data sets we already knew were problematic (e.g., North Pier). Confirmation sampling at a given survey unit will be performed after Phase II evaluation on that survey unit. That is the reason we chose to prepare a global SAP identifying sampling methods and procedures and then prepare Task Specific Plans that address individual survey units that will be resampled.

   EPA Clarification: "Soil columns" means borings (e.g. Geoprobe) down to excavation depth to confirm at 6 inch intervals that concentrations meet the release criteria. Phase 2 sampling results could reveal new anomalies not identified in the statistical tests in Phase 1. In that situation, it may be useful to analyze outliers further using Phase 1 approaches. We will work

together with the Navy to refine the sampling plan through the workgroup process already established.

Thank you for the clarification, see response to EPA general comment #4.

4. Page 1, Section 2 Data Evaluation Plan, Phase 1 Evaluation, Bullet 1: Please add release criteria for decisions, e.g. radionuclides and concentrations.

   The release criteria are contained in Table 1 of the Action Memo (Tetra Tech EC, 2006) and has been referenced in the Tech Memo.

5. Page 1, Section 2 Data Evaluation Plan, Phase 1 Evaluation, Bullet 3
   a. Sub-bullet 1: Statistical tests need to be better defined. Also, please state how the outliers will be calculated and presented. i.e. Will the data for each survey unit be presented with a "Box and Whisker Plot?"
   b. Sub-bullet 2: We suggest the contractor provide example calculations for how this will be performed.

   The Tech Memo has been updated to better define the decision rules and statistical tests. The calculations to identify outliers are based on standard statistical quantities (e.g., mean plus 3 standard deviations) or simple comparisons of values (e.g., result greater than release criteria). The format for final presentation of the results is being developed and will be provided to the Tiger Team for review.

   The response does not address the comment. The response does not address outliers or discuss issues with identifying outliers, given the large size of the data set. For example, outliers that may be identified in an area or survey unit (SU) data set (e.g., 20 to 200 samples) would not be identified if the entire data set is used. Please provide more information about the process for identifying outliers.

   Based on the results of the initial statistical tests, the effectiveness of certain tests in identifying data manipulation is being revisited. For example, identifying outliers has not proven to be an effective method and will not be used to identify evidence of data manipulation. The Tech Memo will be updated to reflect and focus on the most useful tests and methods.

6. Page 2, Section 2 Data Evaluation Plan, Phase 2 Evaluation, bullet 6: A certain percentage of additional random samples should be collected when any anomalous outliers are investigated by confirmation sampling and compared with previous data sets.

   The selection of number and location of samples will be identified in the location/task specific addenda, see response to EPA general comment #2.

7. Page 3, Section 3 Data Quality Objectives, Step 1 Identify the Problem, Identify Decision Making Method: Some additional field work should be performed for anomalous data and identified outliers for both Phase 1 and Phase 2.

Navy Response: Phase I consists of the statistical tests. Phase II is the evaluation of the statistical tests and evaluation of other data as needed. Sampling will be performed if falsification is indicated at the conclusion of Phase II.

EPA Clarification: Phase II could also identify outliers not found in Phase I; therefore, additional field work may also be warranted if those are found. We will work together with the Navy to refine the sampling plan through the workgroup process already established.

See responses to EPA general comments #2 and 4.

8. Page 3, Section 3 Data Quality Objectives, Step 2 Identify the Decision, Bullet 1: Add release criteria decision radionuclides and concentrations. Based on release criteria alone, will outliers be identified that have an uncertainty that is a percentage of the release criteria for confirmation sampling?

See response to EPA specific comment #4 concerning release criteria. A decision rule based on uncertainty is directed towards identifying data quality issues and not data falsification, and is beyond the scope of this project. If it is concluded that data has not been manipulated and is below the release criteria, the data can be used to support decisions for transfer of the real property.

The response partially addresses the comment. The last sentence of the response indicates that based on the statistical tests, it is possible to conclude that data has not been manipulated. However, during Tiger Team meetings, it has been stated that falsification of only a few sample results in a SU cannot be identified by statistical tests. Since falsification could have included only a few samples in a data set and this cannot be identified by the statistical tests, it should not be concluded that the data alone can be used to support decisions for transfer of the real property. As EPA has previously stated, some sampling (e.g., in areas exceeding the NCP risk range prior to remediation) will likely be required. Please delete statements about transfer based only on statistical tests from the Revised Draft Radiological Data Evaluation Plan, Former Hunter's Point Naval Shipyard (the Evaluation Plan).

The statements in the plan are that data may be used to "support decisions regarding transfer of property at HPNS". The statistical tests on the data are not the only line of evidence to support decision making for transfer of property. The other lines of evidence (including sampling as agreed upon by the Tiger Team) are discussed in the Phase 2 section.

9. Page 4, Section 3 Data Quality Objectives, Step 3, Phase 2:
    a. Bullet 2: What percentage of archived samples will undergo a "Physical Review?" What happens if/when any archived samples do not have the same physical appearance of other samples in the survey unit or if some samples are missing?

    b. Bullet 5: Random confirmation sampling should be performed in addition to investigating survey locations with outliers and compared with previous results.

Navy Response: Please clarify the rationale for additional sampling if no evidence of falsification is found.

EPA Clarification: See answer above to General Comment #3.

See response to EPA general comment #3.

10. Page 4, Section 3 Data Quality Objectives, Step 4: A list of prioritized survey units should be provided to the contractor in the order that the datasets should be investigated/assessed and confirmation sampling.

    The Navy has prioritized some sampling sites that can be identified based on former worker allegations of data manipulation (e.g., Building 351A Crawl Space). Once summarized, the results of the statistical evaluations will be provided to the Tiger Team for review. The results will include a list of survey units and dates with statistical differences and can be prioritized for further evaluation and investigation.

11. Page 4, Section 3 Data Quality Objectives, Step 5: The first paragraph should be revised to state that a data quality assessment will be required to ensure the data are usable for decision making in the data evaluation process. Further, it is not clear that it can be concluded that the data can be used to support decisions for transfer without confirmation sampling. Confirmation sampling should be done because these decisions should not be based solely on statistical evaluation. For example, it was stated during the January 5, 2017, conference call that it is not possible to identify whether only a few samples from an 18-sample set were falsified as the evaluation outlined in the Plan only can identify whether an entire 18-sample set were falsified. Because falsification could have been done to resolve apparent hot spots in a data set, it is possible that data for one to three samples in a data set were falsified. Since this cannot be identified using the proposed statistical analyses, confirmation sampling will be necessary. Please revise Step 5 to include a data quality assessment and confirmation sampling.

    Navy Response: This statement, "Further, it is not clear that it can be concluded that the data can be used to support decisions for transfer without confirmation sampling," seems to indicate that no property transfer can occur without additional confirmation sampling – in other words, all survey units must be resampled. This comment is not consistent with EPA's letter to the Navy on 14 December 2016 stating, "Re-sample in priority areas of uncertainty, especially in areas of greatest concern based on health risk" and "Where allegations have been made regarding specific locations on the site, research site records and, where potential health risk is uncertain, sampling and/or scanning should be conducted in those areas." .

    EPA Clarification: EPA is not recommending that all survey units be resampled. However, it would be prudent, by the time the overall assessment is concluded, to ensure that some resampling has occurred in each parcel. Our October, 2016, written recommendations already identified some categories to resample. We will work together with the Navy to refine the sampling plan through the workgroup process already established.

    See response to EPA general comment #2 and specific comment #6.

12. Page 4, Section 3 Data Quality Objectives, Step 5 Develop a Decision Rule, Phase 1 Decision Rules:
    a. All data sets may not have a normal distribution (e.g. from individual survey units. For efficiency, in the case of non-normal distributions, additional analysis may not be helpful or necessary.

b. Summary Statistics, bullet 3: Likewise, radiological results reported as detections, but which include a counting or total propagated uncertainty (TPU) error that is greater than the reported result, should be considered un-reliable and flagged for investigation in Phase 2.

Navy Response: This comment is inconsistent with the opinion of the Navy and others on the team (e.g., CDPH) that are in favor of using professional judgment.

EPA Clarification: As we said on the phone, in situations where review of visual displays are the method of identifying potential anomalies, then EPA can review these concurrently through webinars, in-person meetings, emailed images, or other approaches. We will work together with the Navy to refine the sampling plan through the workgroup process already established.

See responses to EPA general comments #1 and 2. Results of data evaluation will continue to be shared prior to and for discussion on the bi-weekly technical calls. Additional working meetings can be scheduled as needed.

The response does not address the comment. It appears that there may still be some misunderstandings about the intent of the comment. The Evaluation Plan should discuss how data sets that do not have a normal distribution will be handled for tests that assume a normal distribution. In addition, data with total propagated uncertainty (TPU) error that exceeds the reported results should be considered for Phase 2 investigation. Please discuss how data sets that do not have a normal distribution will be handled and ensure that data with TPU error that exceeds the reported results be considered during Phase 2.

All of the statistical tests are non-parametric and do not assume normal distribution. A total TPU greater than the reported result is a data quality issue and data quality was assessed during previous document reviews and approvals and is not the focus of this evaluation.

13. Page 5, Section 3 Data Quality Objectives, Step 5, Phase 1 Decision Rules, Bullet 2 Histograms
    a. Sub-bullets 1 and 2: Please describe how outliers will be flagged.
    b. Sub-bullet 4: We suggest striking "professional judgement" and adding how outliers will be mathematically identified.

Outliers will not be flagged solely based on the histograms or other graphical representations of data. The histograms provide visual patterns that humans can better identify versus a computer program (e.g., professional judgment). If an unusual distribution is identified, it will be used as one of several lines of evidence for identifying the parcel or survey unit for further evaluation and investigation. Therefore, we do not recommend removing professional judgment from interpretation of statistical analyses.

14. Page 5, Section 3 Data Quality Objectives, Step 5, Phase 1 Decision Rules, Bullet 3 Normal probability plots:
    a. Sub-bullet 1: This may not be possible since Ra-226 and eRa-226 are not identified separately. What is the definition for identifying outliers? 1st/3rd Quantile ±1.5(IQR),1st/3rd Quantile ±3(IQR)?
    b. Sub-bullet 2: How will this be presented graphically? We suggest that the contractor present a "test case" from beginning of evaluation to the end and include decision points.

See response to EPA specific comments #2 and #13. The format for final presentation of the results is being developed and will be provided to the Tiger Team for review. Results of data evaluation will continue to be shared prior to and for discussion on the bi-weekly technical calls. Additional working meetings can be scheduled as needed.

The responses do not address the comments. Please review the original comments and explain how outliers will be identified.

Based on the results of the initial statistical tests, the effectiveness of certain tests in identifying data manipulation is being revisited. For example, identifying outliers has not proven to be an effective method and will not be used to identify evidence of data manipulation. The Tech Memo will be updated to reflect and focus on the most useful tests and methods.

15. Page 5, Section 3 Data Quality Objectives, Step 5, Phase 1 Decision Rules, Bullet 4 Posting plots: How will outliers in the Final Status Survey results be defined? Will they be based on a survey unit or all parcel data? It is not clear that it will be possible to identify falsification from posting plots if a sample was moved to a location that was less likely to be contaminated within the triangular systematic survey grids.

    Posting plots are expected to be one of several lines of evidence for identifying parcels or survey units for further evaluation and investigation.

    The response does not address the comment. It is unclear how posting plots can be used to identify the case where samples were not collected from the randomly selected area, but instead were collected from an area that was known to be less contaminated. Please address this issue.

    Posting plots do not identify cases where someone has collected a sample elsewhere and gives coordinates that place it within a SU at the specified random start, triangular grid sample location. Posting plots can give indications of clustered areas with elevated contamination during repeated characterization, remedial action, and/or final status phases of sampling that has a final sample result much different than previous iterations. It is only one line of evidence among many that may aid the investigation. However, because x,y coordinates are not available electronically for most sample locations, the usefulness of posting plots is limited.

16. Page 5, Section 3 Data Quality Objectives, Step 5, Phase 1 Decision Rules, Bullet 5 Time Series Plots, sub-bullet 1: Will a plot be generated for all radionuclides or just the radionuclides identified for clearance in the ROD? i.e. K-40 and other natural nuclide progeny.

    Time-series plots are being generated for all radionuclides.

17. Page 6, Section 3 Data Quality Objectives, Step 5, Phase 1 Decision Rules, Bullet 9 Two-Dimensional Paired Kolmogorov-Smirnov (Peacock) Test: The proposal to evaluate equilibrium comparisons should consider the following:
    a. The Two-Dimensional Paired Kolmogorov-Smirnov (Peacock) Test would be most effective when a reliable data set from Hunters Point Shipyard is used as the control data set, but this is not proposed. It is unclear how this test can be effective without

a reliable control data set for comparison, particularly since there are questions about the reliability of the entire data set that is under evaluation. Please revise the Plan to include use of a reliable data set from Hunters Point Shipyard or explain how this test can be effective without one.

In terms of falsification, the test as proposed in the Tech Memo will be effective to show the relationship between the parent and the daughter radionuclides demonstrated in all other survey units within the same parcel and identify the survey units that stand out as different for further investigation. Each survey unit is compared against the remainder of the parcel data set as a control. For example, it will show whether someone may have altered values for one radionuclide (e.g., radium) but not another (e.g., bismuth). A reliable control data set, if it exists, has inherent difficulties in gaining consensus and as being representative for application to a large data set collected over many years and therefore may or may not improve the effectiveness of the test in identifying falsified data.

b. Offsite laboratory results may not provide a separation of Ra-226 activity concentration from eRa-226 concentration whenever an ingrowth method was used. The narrative of the data packages may be the only place where this could be found, but the narrative may not include which energy lines were averaged after ingrowth to quantify Ra-226 (Pb-214, Bi-214, or both). For the Ra-226 results generated from off-site laboratories, the data should be examined to determine if the sample was placed in a sealed container and stored for 21 days prior to analysis. If the 21-day ingrowth was not completed, an evaluation of how this was accounted for in the reporting of the final results should be investigated. Please ensure that the calculations and data sets used in such evaluations are presented to the team members to ensure all parties are in agreement with the approach. The Contractor will need to have copies of analysis libraries to see which energy lines are used for quantifying Ra-226. Ingrowth method may use Pb-214/Bi-214 lines; hence no Pb-214/Bi-214 results.

See response to EPA specific comment #2.

The response does not address the comment; simply referring to "the response to EPA Specific Comment #2" is insufficient. Please revise the response to specifically address the original comment.

The off-site laboratory results are being investigated to identify evidence that a replacement sample was sent to the laboratory for analysis, the actual analysis performed by the offsite laboratory is not being evaluated.

c. It is unclear why Th-232 results from site samples will be compared to Pb-212 and Bismuth 212 (Bi-212) results rather than using results from Actinium-228 (Ac-228), with a 6.1-hour half-life to evaluate the equilibrium status. Using the Pb-212 radionuclide to evaluate the equilibrium status would require a forty-year ingrowth time to establish secular equilibrium. If a ratio of Th-232 concentrations to Pb-212 concentrations is established to complete the comparison, information regarding

the assumptions and correction factors applied will need to be provided. Please provide additional information regarding how this evaluation will be conducted. A comparison of Th-232 by alpha spec should be done with all daughter progeny by gamma spec (Ac-228, Pb-212, Bi-212 and Tl-208).

The format for final presentation of the results is being developed and will be provided to the Tiger Team for review. Results of data evaluation will continue to be shared prior to and for discussion on the bi-weekly technical calls.

d. Please specify the statistical confidence level associated with a p value of 0.05.

The p value of 0.05 is standard and 100(1-p)% gives a confidence level of 95%.

18. Page 6, 3 Data Quality Objectives, Step 5 Develop a Decision Rule, Phase 1 Decision Rules, Bullet 9: For the Benford's Law Test, most first digits in the dataset would be expected to be 0 or 1 for many of the radionuclides, whether detected or falsified. For example, it is possible that a result of 3.4 picoCuries per gram (pCi/g) was modified to 1.4 pCi/g or even 0.4 pCi/g. It does not appear that this test or any of the statistical tests would detect such a change to an individual result. The second, third, and fourth integer tests may produce more meaningful results.

Agreed, the example provided is not likely to show up on the Benford's Test and non-statistical approaches (e.g., comparing lab results to the report) will be one of several lines of evidence used to identify this type of falsification. The only digits that are being analyzed in the Benford's Test are the first or the first and the second non-zero digits. This has been clarified in the Tech Memo.

The response partially addresses the comment. It is unclear how the statistical tests can be used to identify falsification where the first digit only was changed to 0 or 1, as discussed in the original comment. In addition, please explain how this type of falsification can be identified if only one or two samples in a survey unit have modified results.

The Benford's Test looks at the first non-zero digit or the first two non-zero digits; therefore, a first digit of zero is not an option. Identifying only one or two samples that have been falsified will not likely be identified using the Benford's Test. To identify potential falsification for individual results, logic tests (e.g., documentation reviews) will be used as the primary indicator.

19. Page 6, 3 Data Quality Objectives, Step 5 Develop a Decision Rule, Phase 1 Decision Rules, Bullet 10 General Logic: How will sample locations within the survey unit be selected?

See response to EPA specific comment #6.

20. Page 6, 3 Data Quality Objectives, Step 5 Develop a Decision Rule, Phase 2 Decision Rules: During Phase 2, recounting selected archived samples that are determined to be consistent with the survey unit soil type should be done to validate the data set and conclusions from Phase 1. For example, it is unclear if there are cases where data, but not samples, were manipulated. Further, analyzing selected archived samples may help in evaluation of cases where only a few samples in an 18-sample data set were falsified. In addition, validated laboratory results with counting uncertainty/TPU and validation flags need to be evaluated. It is our understanding that

the NIRIS database was not populated directly from laboratory EDDs. Where validated data is not available, then for efficiency, further analysis may not be helpful or necessary.

See response to EPA general comment #1. Recounting archived samples is one line of evidence that may be considered to identify evidence of possible falsification.

21. In the future, as the plans evolve, a decision tree should be provided so that it is clear how decisions will be made. For example, a decision tree would graphically illustrate when survey units/samples would be recommended for Phase 2.

    Agree, a decision tree will be developed as part of Phase 2.

22. In the future, as the plans evolve, please address buildings. More specifically, on December 7, 2016, EPA gave the Navy an updated version of its detailed technical recommendations. On page 7, Recommendation 7 stated "The Navy should also address the elevated levels found in Buildings 271 and 406. . . . The Navy should address the implications of these findings for these and other buildings." The calculations shown in the attachment raise questions about several areas of previous work conducted by Tetra Tech EC in buildings. Based on the information we have at this time, EPA would recommend some rescanning of buildings at the site. We therefore support the recommendations for rescanning from the State of California Department of Public Health (CDPH) letter in November, 2016. We understand that the Navy has chosen to evaluate buildings on a longer time frame and to focus on evaluating soil data first. When in the future more resources are available to study the buildings issue in more depth, we will look forward to revisiting this concern and reviewing any new information that may help us understand better any potential risks to public health and the environment related to buildings.

    Buildings will be considered in future plans, evaluations, and investigations.

ADDITIONAL COMMENTS

1. Section 2 (Data Evaluation Plan), Step 1 Evaluation: There are several concerns associated with the Step 1 Evaluation. The first bullet in the Phase I Evaluation states, "A statistically determined number of reports and individual data results from written reports (RACRs/SUPRs) will be compared with electronic results of NIRIS and Tetra Tech EC databases to identify potential manipulation. This analysis is biased toward reports from sites with areas with the greatest exceedances of cleanup levels prior to remediation being selected more often than reports from sites with less exceedances prior to remediation. Data will also be compared to laboratory reports once they are obtained. It should be noted that the Navy RASO [Radiological Affairs Service Office] reviewed original laboratory data reports so falsification of the laboratory results database, if it occurred, is not likely to have affected decision-making."

    a. While it is understood that a statistical number of data reports were checked to determine the accuracy of results between the NIRIS and Tetra Tech EC databases, it is not clear how the process to identify data outliers will proceed given that there are multiple databases that have been determined to contain discrepancies. Additionally, as noted in EPA General Comments 1 and 2, use of data should be based on a determination that it is sufficiently reliable. Please provide a more detailed data evaluation plan that specifies which databases and/or hard copy data reports will or are being accessed to perform data anomaly evaluations. Please

ensure that the data evaluation plan also discusses how the sources of data were determined to be comparable for meaningful statistical analysis and evaluation. This should include listing the source of the data sets, the timeframe the data was collected, whether the counting and/or total propagated uncertainty associated with such results was reported and evaluated, and any other pertinent information related to the quality/reliability of such data sets.

The initial decisions to transfer property at HPNS were based on available written reports (RACRs, SUPRs, FSSRs); therefore, these reports represent the data sets to be evaluated. The electronic database is being compared with the written documents to ensure the database accurately reflects what was provided in the reports. The list of reports and individual samples included in the comparison will be documented in a final report that will be provided for stakeholder review.

b. The statements in this first bulleted item indicate that because RASO reviewed original laboratory data reports, falsification of the database, if it occurred, is not likely to have affected decision-making. While this statement may have originally been intended to be limited to apply specifically to the narrow concern about discrepancies between laboratory reports and the NIRIS database, however, review of laboratory data reports does not preclude the possibility that samples were tampered with prior to analysis. Therefore, this statement could be misleading. In fact, the Nuclear Regulatory Commission concluded in its enforcement action that Tetra Tech EC, Inc., "deliberately falsified soil sample records by taking soil samples from areas not designated as part of the target area and by completing forms with inaccurate information on a number of occasions in late 2011 through mid-2012." (Source: NRC press release, October, 2017) In addition, on the January 31, 2017, conference call, the independent contractor stated that the NIRIS database did not match the lab reports in some instances. Please either delete this or re-word it to also communicate the need to evaluate data to determine whether samples were substituted and/or other such means were used to falsify results.

The statement will be deleted.

2. Section 2 (Data Evaluation Plan), Step 1 – Identify the Problem: The Navy is identified as the primary decision maker. However, the data evaluation plan does not acknowledge that the regulatory stakeholders fulfill oversight and regulatory compliance assessment roles, and therefore are also part of the decision making team. Please revise this statement to acknowledge the collaborative process for ensuring cleanup at the HPNS complies with the CERCLA cleanup process, the Federal Facility Agreement, as well as compliance with all State and Federal environmental, radiological, and safety regulations.

   The statement will be revised to reflect the collaborative process for ensuring cleanup at the HPNS complies with the CERCLA cleanup process, the Federal Facility Agreement, as well as compliance with all State and Federal environmental, radiological, and safety regulations.

3. Radionuclides of concern: Thank you for the January 31, 2017, presentation about focusing strategically on certain radionuclides of concern. Overall at the HPNS, based on the history of the site and data collected from multiple contractors, Ra-226 has the greatest likelihood for the highest activity with health implications prior to remediation. For most of the HPNS, EPA

supports the approach proposed by the third party independent consultant. In some specific areas where particular radionuclides of concern other than Ra-226 and Cs-137 have been measured, please include those additional radionuclides in the statistical analyses. Some of these specific areas were identified on the January 31, 2017, conference call.

Here are examples where we recommend further evaluation:

1. Strontium 90 (Sr-90) and Cesium 137 (Cs-137) in Parcel E 500 series buildings and the Building 707 triangle areas where NRDL operations occurred.

2. Sr-90 in luminescent deck markers or other radiological commodities in areas where Sr-90 devices have been found.

3. Sr-90, Cs-137, and Plutonium 239 (Pu-239) in a vault together in the Parcel E 500 series area.

4. Thorium 232 (Th-232) in Building 130 gravel in Parcel B.

5. Americium 241 (Am-241) in the vicinity of the building used for Geotechnical testing.

In addition, for the record, here are two situations that EPA would not consider a priority for further evaluation:

Cobalt 60 (Co-60): The Navy ceased Shipyard operations in 1974, 42 years ago. The half-life of Co-60 is 5.26 yrs. After seven to ten half-lives (i.e., 37 to 53 years), remaining radiological activity would be at levels similar to background. Therefore, Co-60 is not a priority health and safety concern, and any Co-60 sampling conducted would not be a helpful indicator of potential prior falsification.

Europium: Europium was found in the Gun Mole Pier in Parcel D-1, but that work was done by Shaw, and not Tetra Tech EC, so it is not a priority for this evaluation.

Agree, the radionuclides of concern (ROCs) listed in the individual reports (RACRs, SUPRs, FSSRs) will be the ROCs for data evaluation. The Tech Memo will be updated to focus on the ROCs. However, if there is insufficient data to support statistical data analysis, no statistical test will be performed.

DTSC COMMENTS (January 11, 2017)

1. Step 3 last bullet: Please include that 10-20% of the samples collected will be split and submitted to CDPH for analysis by the state.

   Samples will be split and shared as agreed to and in accordance with the UFP-SAP.

2. Step 5 Paragraph 1: This does not address the issue of collecting supplemental samples to replace manipulated data (if possible) and then determining if the new data results allow for the property transfer decisions.

   The general statement referenced is not intended to convey every effort needed to "support the decisions for transfer of the real property". Additional sampling may be one of several lines of evidence needed to support the decisions.

3. Step 5 Phase 2 Decision Rules: Are you saying compare the data TT EC put in the database with that which they put in the hard copy reports? If falsified data was used in the database it was most likely portrayed in the report that followed. I'm not sure this would be valid and would

consider deleting. Please clarify/revise. MIL-STD-105E was canceled. The current Notice of Cancellation (Notice 3) recommends that future acquisitions refer to: MIL-STD-1916, "DoD Preferred Methods for Acceptance of Product" (available online at http://assist.daps.dla.mil), or ANSI/ASQ Z1.4, "Sampling Procedures and Tables for Inspection by Attributes" (available online at www.asq.org).

Yes, the Tetra Tech EC database is being compared to the NIRIS database and the available hard copy lab reports. The reference has been updated as suggested.

4. Step 5 Phase 2 Decision Rules: Do you mean collected a long period of time after completed?

   The statement is intended to indicate post-remediation/FSS samples and this has been clarified in the text.

5. If possible, add a schedule.

   A project schedule will be developed, reviewed, and approved by the Tiger Team separately.

DTSC COMMENTS (February 2, 2017)

1. Section 3 - Data Quality Objectives, Step 1, References seven parcels at the property while Step 4 in this same section references 10 parcels. It is unclear how many and which parcels will have data evaluated. Please clarify in the text of each section.

   Parcels B, C, D, E, E-2, F, G, UC-1, UC-2, and UC-3 are being evaluated and the text will be updated for clarification.

2. Section 2 – Data Evaluation Plan, Phase 2 Evaluation, Bullet 5 has been revised to include sample collection in the UFP-SAP to address "previous allegations and concerns", while in Section 3-Data Quality Objectives, Phase 2 Decisions Rules, the last bullet references EPA recommendations and concerns. Both sections appear to discuss confirmatory sampling and therefore, should be consistent.

   a. The reference to include sample collection to address previous allegations and concerns should be added to Section 3 - Data Quality Objectives, Step 2, last paragraph.

   The text will be updated as recommended.

CDPH Comments (January 11 and February 2, 2017)

EMB will determine the percentage of number of split samples collected for confirmation.

1. Page 1, Section 2 Data Evaluation, Plan Phase 1 Evaluation, Bullet Point One: Which statistical methods or other evaluation will be used to identify discrepancies in data that was manually entered into the database? This information needs to be included in the plan.

   A statistically determined number of reports and individual data results from written reports (RACRs/SUPRs) were compared with electronic results of NIRIS and Tetra Tech EC databases to identify potential manipulation. This analysis was biased toward reports from sites with areas with the greatest exceedances of cleanup levels prior to remediation being selected more often than reports from sites with less exceedances prior to remediation. Data will also be compared to laboratory reports once they are obtained. The Tech Memo has been updated to

reflect this. Please note that RASO reviewed original laboratory data reports so falsification of the laboratory results database, if it occurred, is not likely to have affected decision-making.

2. Page 2, Phase 2 Evaluation (still under development), Bullet Point One: Please indicate that these records are for soil data only.

   This has been clarified in the Tech Memo.

3. Page 3, Step 2 – Identify the Decision: Which statistical methods or other evaluation will be used to identify samples collected from a biased location suggesting falsification? This information needs to be included in the plan.

   Specific falsification allegations will be evaluated by reviewing dates of sample collection, chains of custodies, log book entries, inspection of sample archives, and draft and final reports as available. This has been added to the Tech Memo.

   Please identify the statistical methods can be possibly used.

   The statistical tests are listed in Step 3 of the plan. However, based on the results of the initial statistical tests, the effectiveness of certain tests in identifying data manipulation is being revisited. For example, identifying outliers has not proven to be an effective method and will not be used to identify evidence of data manipulation. The Tech Memo will be updated to reflect and focus on the most useful tests and methods.

4. Page 5, First Bullet: The initial lab method used for the analyses was different for the on and offsite laboratories. It wasn't until around 2012 that the laboratories used the same methods. How may this affect the evaluation and comparison of the data and how will it be addressed if necessary?

   Agreed, the available data, regardless of method, will be used in the statistical evaluations; however, differences identified will be further evaluated and investigated to determine the cause, and the specific analyses in which it can be used (e.g., parent-progeny equilibrium).

5. Figure 2: Building Scans are explicitly mentioned as part of the Phase 2 in Figure 2. The main document does not outline this effort. Please modify the document to reflect the concerns and evaluation indicated in Figure 2. As a reminder, we have prepared a memorandum recommending rescanning of all Class 1 survey units for buildings in Parcels C and E (Attached).

   The Navy is still reviewing the building scan data set, and evaluating potential analyses that may indicate data manipulation. The primary discrepancy in the building scans involved Tetra Tech EC personnel using a faster than agreed-upon scan rate. This matter is being addressed and buildings will be considered in future plans, evaluations, and investigations. The Tiger Team will be provided plan(s) to review and approve.

WATER BOARD COMMENTS (January 11, 2017)

1. One of the allegations is that technicians collected soil samples from clean areas instead of the locations specified in the work plans. This possibility is discussed in Step 2 of the Data Quality Objectives. Please clarify how Phase I of the evaluation will comprehensively and thoroughly

answer the question of whether there is any evidence that soil samples were not representative of the matrix (e.g., not collected from the survey unit soil matrix). What would be expected of soil samples that are representative of the survey unit soil matrix? What possible data could be evaluated and how did the team select the proposed evaluation methods?

Specific falsification allegations will be evaluated by reviewing dates of sample collection, chains of custodies, log book entries, inspection of sample archives, and draft and final reports as available. This has been added to the Tech Memo. To identify soil samples that are not representative of the survey unit soil matrix, geological information will be reviewed from available soil boring and test pit logs and sample archives and compared to samples across the survey unit and to any new samples collected. Development of Phase 2 DQOs is ongoing and soil matrix will be considered.

Soil samples that are representative of a survey unit would be likely to have similar radioisotope activities as nearby survey units and as the pre-remediation samples collected at the survey unit in question. The proposed statistical methods will aid in discerning survey unit soil from soil collected elsewhere.